MODIS DATA SYSTEM STUDY

TEAM PRESENTATION

September 16, 1988

AGENDA

1. An Example of System Sizing: the "Calibrate" Function

2. MIDACS Context Diagrams and Data Dictionary

3. Notes from Conversation with Gene Smith on DHC Operations

4. Action Items

An Example of System Sizing: the "Calibrate" Function

In the early stages of this MODIS Phase-A/B study, our understanding of the MODIS Information, Data, and Control System (MIDACS) functional and performance requirements is necessarily limited to general concepts. As our knowledge of these requirements improves, through interactions with the MODIS Instrument and Science Teams, as well as the EosDIS, we are able to provide an increasingly quantitative specification of the data system. We have previously presented rough order of magnitude estimates of the level-1A, 1B, 2, and 3 MODIS-N and MODIS-T processing throughput requirements in terms of MIPS based on general EosDIS path length guidelines and the assumed data rates for the instruments. Clearly, it is important to refine these estimates as much as possible to accurately specify a viable data system for MODIS at the end of the study.

At this time, there does not exist sufficient information to accomplish this refinement for the higher data levels (e.g., level-3 and level-4). However, for level-1A, 1B, and 2 (particularly -1A and -1B), we can now begin to more clearly specify the performance requirements dictated by the probable (though assumed) MODIS algorithms.

Here we consider one function of the level-1 processing: the counts to radiance calibration. We assume the calibration to be a linear process, involving the multiplication (division) of the detector voltage by a sensitivity and the addition of an offset. The arithmetic expression is of the form $y_i = a_i * x_i + b_i$, where the coefficients a_i and b_i are the calibration "constants," x_i is the detector voltage, y_i the calibrated radiance, and the subscript i denotes the detector number. To evaluate this expression, approximately 10 instructions are needed: get the address of a_i, load a_i, get the address of x_i, load x_i, multiply a_i and x_i, get the address of b_i, load b_i, add b_i to the product of a_i and x_i, get the address of y_i, and store y_i. In addition, about three instructions are required to cycle through the loop: increment the counter, compare, and branch. We thus expect a total of 13 instructions for each calibration. (In reality, a_i and b_i are often temperature-dependent, requiring several additional instructions for the evaluation of each dependence.)

The raw MODIS data rate, in the absence of oversampling, is on the order of 10^7 bits per second, and about 10^6 observations per second. Multiplying the number of observations per second with the number of instructions per calibration, we find that 13 MIPS are required to calibrate the MODIS data with the processor fully utilized. With a processor utilization of 70%, and neglecting optimization, this requirement reaches 19 MIPS.

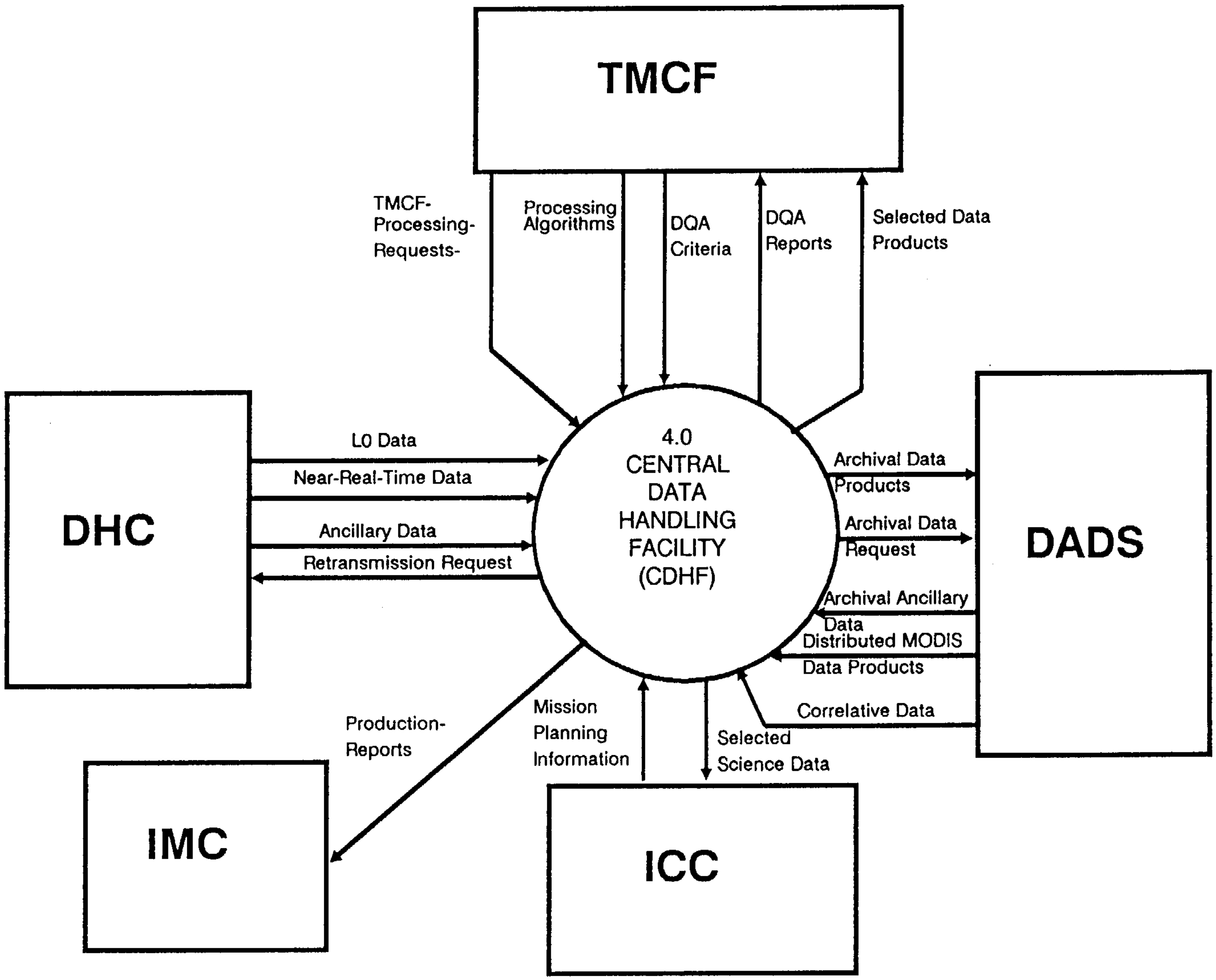
To consider the magnitude of this requirement, the IBM 308X series has a capacity of 10 to 27 MIPS. In the NSESCC, the IBM 3081 is rated at 14.5 MIPS. The implication of the previous paragraph is that the 3081, operating at nearly 100% of its capacity, would need to operate 24 hours a day to simply apply a one-time linear calibration to the MODIS raw data.

The above exercise provides a useful tool for sizing specific aspects of the MIDACS. It is still subject to further refinement. We have tested this analysis by simulating the calibration of 1000 scan positions of the MODIS-T instrument, each position yielding 64^2 observations (64 spectral channels and 64 spatial channels). The algorithm is provided in the Appendix. On the NSESCC 3081, 10 seconds were required for all the calibrations, yielding an effective rate of 5.3 MIPS (4096 channels x 1000 observations x 13 operations/channel / 10). The example behaves as if 35 "instructions" are required for each calibration, rather than the estimated 13. As a result, the IBM 3081 used as an example here, operating 24 hours a day, would not be capable of calibrating the MODIS data.

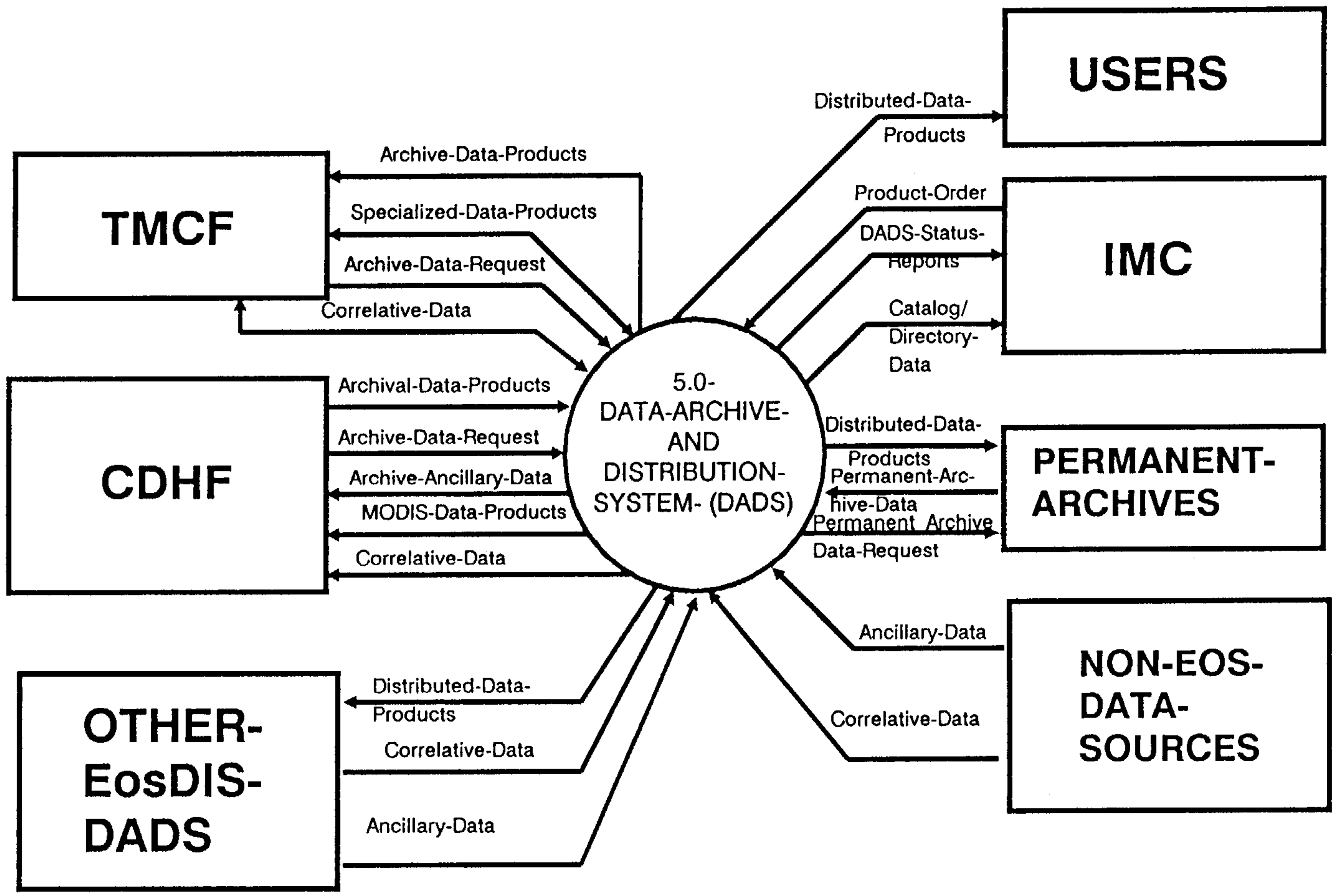
We have incurred an error factor of about 2.5 in our estimate, with the calibrations being applied considerably slower than our estimate. While the estimate is within an order of magnitude, the uncertainty is still larger than desireable. One factor at work here is the fact that some instructions consume more clock cycles than others. One implication of the above discrepancy is the need for careful evaluation of "pilot" algorithms during the process of sizing the system.

Appendix

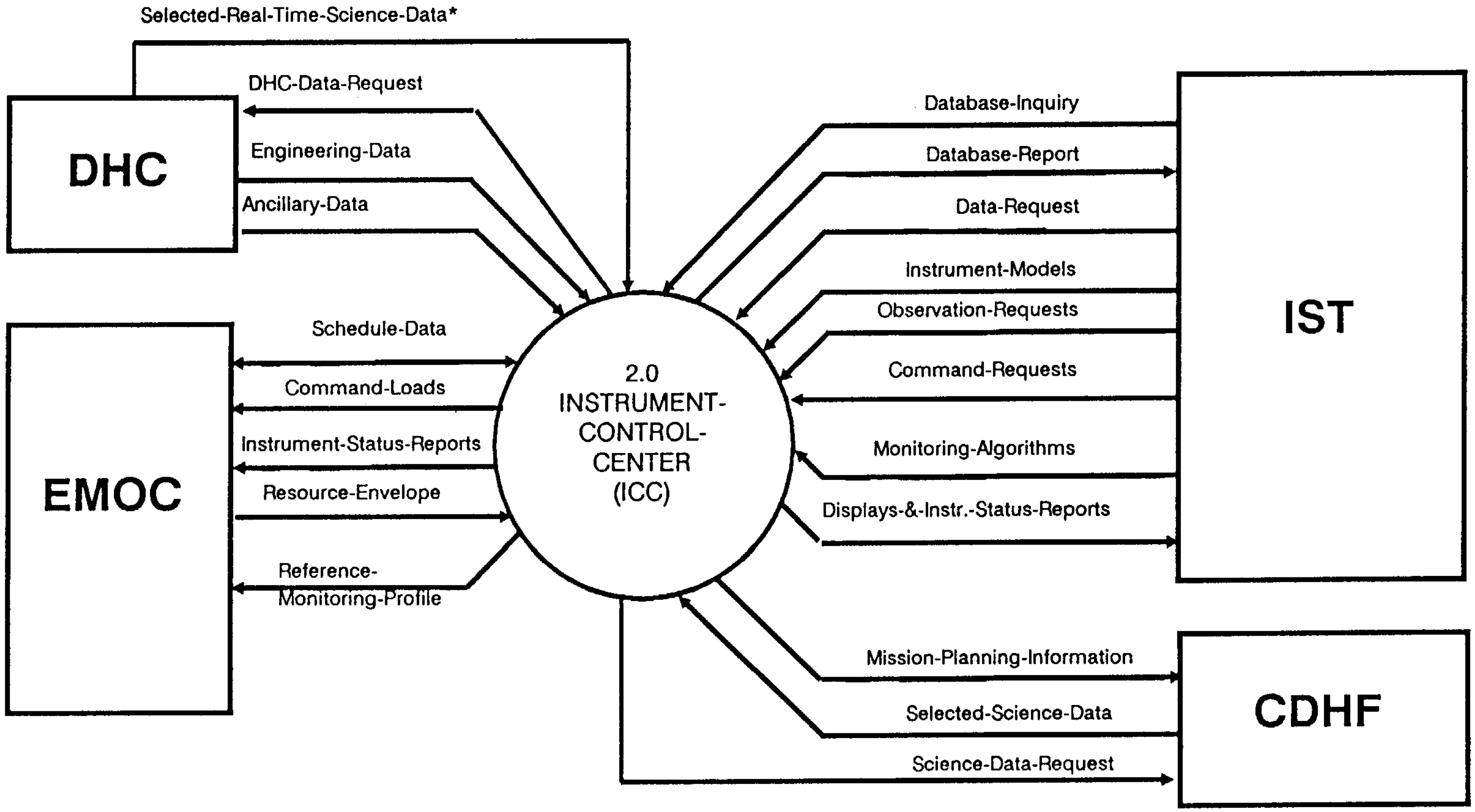
```
      PROGRAM CALIBR

C----------------------------------------------------------------------
C
C   THIS PROGRAM TESTS THE SPEED OF A GENERAL
C   LINEAR COUNT-TO-RADIANCE CALIBRATION.
C
C----------------------------------------------------------------------

      REAL A(4096),B(4096),X(4096),Y(4096)

C----------------------------------------------------------------------
C   INITIALIZE DATA (64 SPECTRAL CHANNELS, 64 SPATIAL CHANNELS)
C----------------------------------------------------------------------


      NTEST=1000
      DO 80 I=1,4096
       A(I) = I
       B(I) = I
       X(I) = I
 80   CONTINUE

      CALL REMTIM(ICPU)
      PRINT *,ICPU

C--------------------------------------------------------------------
C   REPEAT LOOP 1000 TIMES FOR TEST PRECISION
C----------------------------------------------------------------------

      DO 100 J=1,NTEST

C----------------------------------------------------------------------
C   CALIBRATE MODIS-T DATA
C----------------------------------------------------------------------

        DO 100 I=1,4096
          Y(I) = A(I)*X(I)+B(I)
 100  CONTINUE

      CALL REMTIM(ICPU)
      PRINT *,ICPU

      STOP
      END
```



CDHF Context Diagram

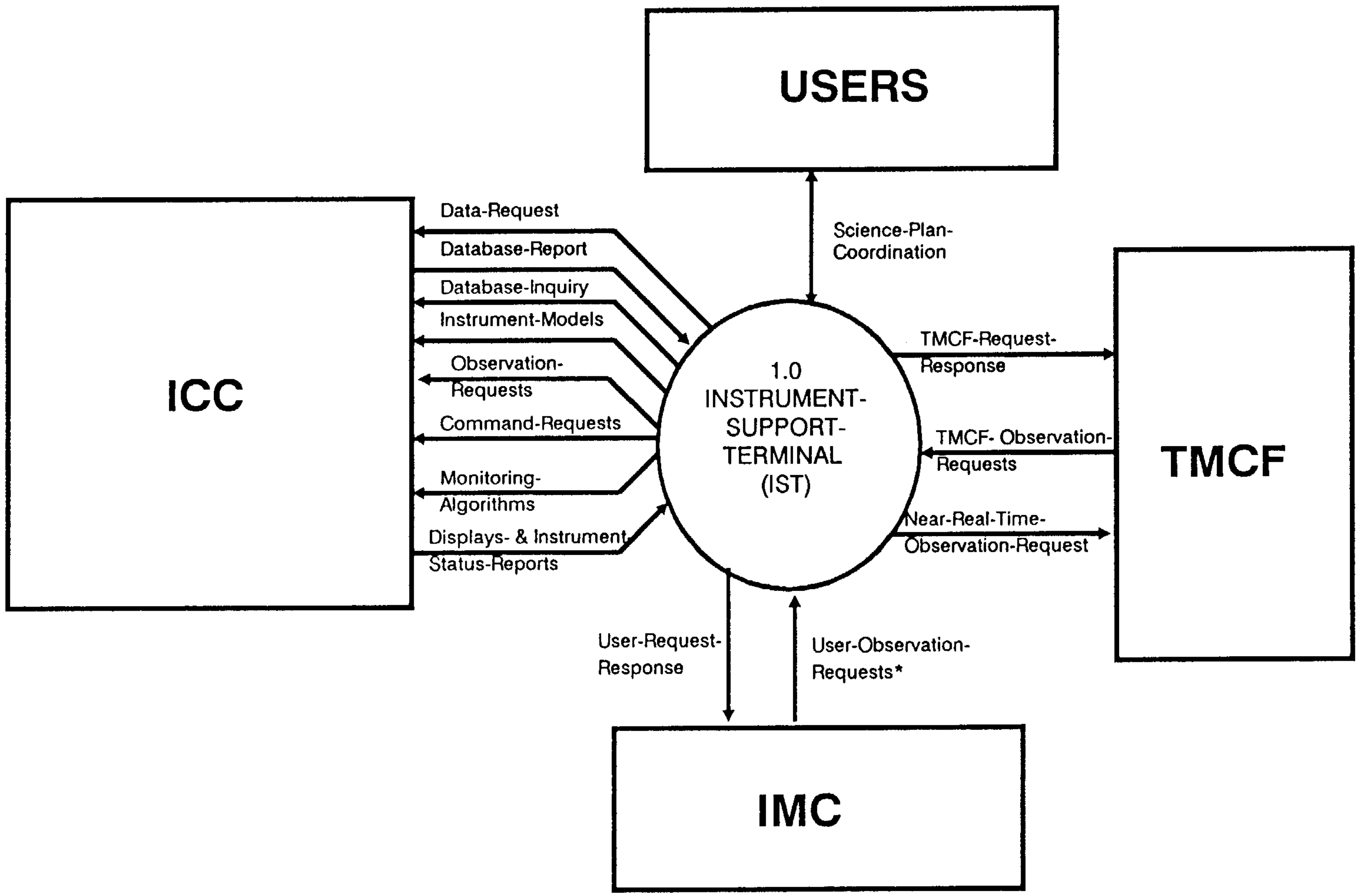


DADS Context Diagram



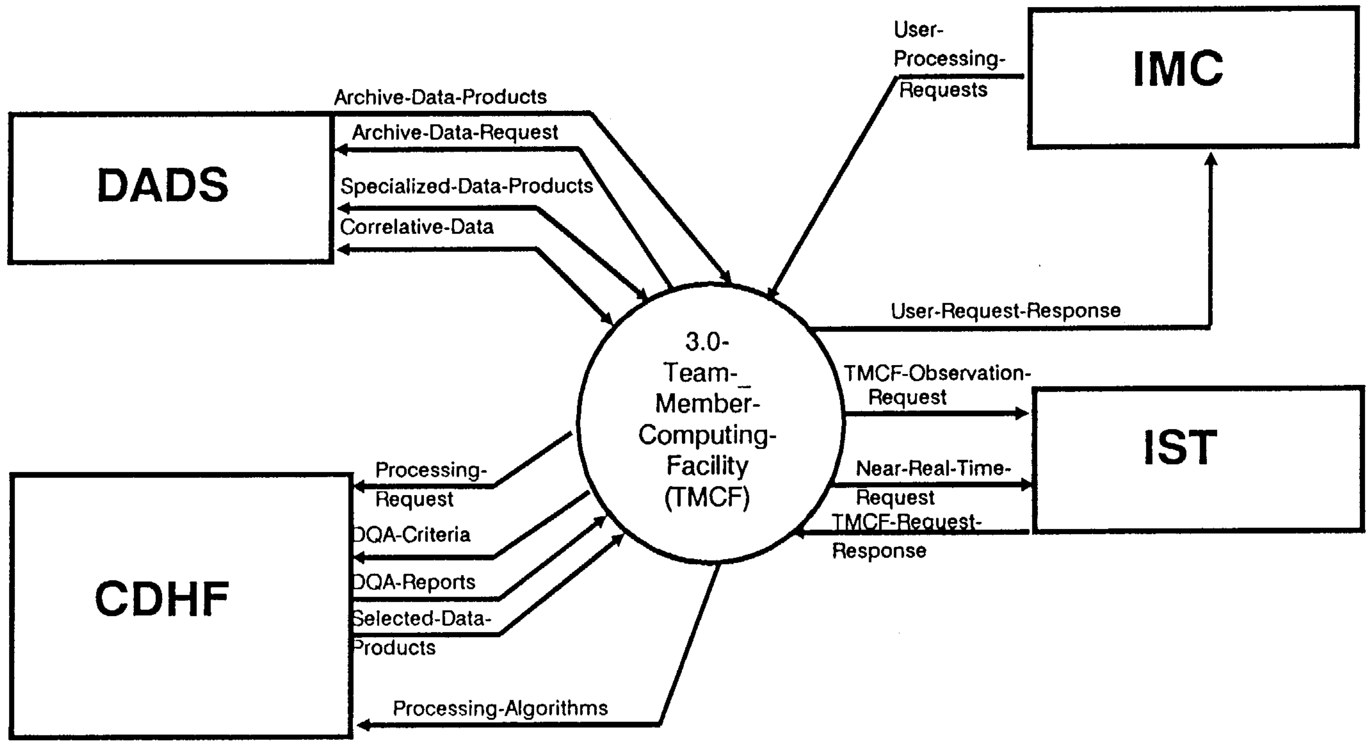
* Issue under discussion

ICC Context Diagram



* Issue under discussion

IST Context Diagram



TMCF Context Diagram

DATA DICTIONARY

Ancillary-Data	=	*Data other than MODIS-Instrument-Data required to perform MODIS data processing. [They include orbit data attitude data, time information, spacecraft or platform engineering data (e.g., pointing information, optics temperature, structure temperature, instrument mounting alignment), calibration data, data quality information,and data from other instruments (e.g., cloud information derived from a second instrument, status of items in a second instrument which could create interference with the instrument data being processed,map data, atmosphere temperature grids).]*
Archival-Data-Product	=	*Any product generated by processing in the CDHF that will be archived for further analysis and/or distribution.*
Archival-Data-Request	=	*Data Requested from any EosDIS DADS except MODIS.*
Catalog/Directory-Data	=	*Listings of data available from the MIDACS DADS listed by platform, instrument, data processing level, algorithm identifier, parameter, time, geographic location, or combination.
Command-Loads	=	*Encoded MODIS instrument command sequences as required by the onboard MODIS instrument control system and constructed so as to affect a specific action; e.g., "HV PWR ON"..*
Command-Request	=	*A command load generated by the IST, verified by and immediately transmitted by the ICC.*
Correlative-Data	=	*Scientific data not from the MODIS instrument used to verify, interpret, or validate MODIS data products.*

DADS-Status-Report	=	*Description of the DADS status, resources utilization, and performance.*
Database-Inquiry	=	*Inquiry of the monitoring database to determine what instrument monitoring reports, data, and analysis are currently available.*
Database-Report		*Report of instrument monitoring functions and availability.*
DHC-Data-Request	=	*Redesignation of packet handling and processing priorities.*
Displays-and-Instrument-Status-Report	=	*A display, (plots, images, list of requested data or status of the instrument or ground system.*
Distributed-Data-Products	=	*MIDACS products routinely archived for potential user access and distributed in response to a product request.*
DQA Criteria	=	*Factors used to assess data quality.*
DQA Reports	=	*Results of routine data quality assessment associated with data receipt and data product operation.*
Engineering-Data	=	*MODIS-Engineering-Data + Platform-Engineering-Data.*
Instrument-Models	=	*Computer compatible mathematical equivalent of the MODIS instrument, used to estimate resource requirements during a modeled operation.*
Instrument-Status-Report	=	*Information on the operating configuration of the MODIS instrument.*
Level-0-Data	=	*MODIS-Instrument-Data at original resolution, time order restored, with duplicates removed.*
Mission-Planning-Information	=	*Information provided by the ICC to the CDHF which may be useful for data production scheduling.*
MODIS-Engineering-Data	=	*Data other than MODIS-Science-Data generated within the MODIS instrument.*

MODIS-Instrument-Data	=	*Data originating within the MODIS instrument.*
	=	MODIS-Science-Data + MODIS-Engineering-Data
MODIS-Science-Data	=	*Unprocessed radiance observations as generated by the MODIS instrument.*
Monitoring-Algorithms	=	*A procedure (or recipe) for transforming information into a different state to accommodate a specific data interpretation.*
Near-Real-Time-Data	=	*MODIS-Instrument-Data designated for Priority Processing.*
Near-Real-Time-Request	=	*Request to handle data in Priority Mode.*
Non-Standard-Products	=	*Products not routinely produced, standard products produced by an alternate algorithm, or combinations of standard products.*
Permanent-Archive-Data	=	*Data retrieved from permanent archival storage.*
Permanent-Archive-Data-Request	=	*Request for data from the permanent archive.*
Platform-Engineering-Data	=	*Data produced by the platform sensors that one used for operating the platform or as ancillary data.*
Priority Processing	=	*Immediate processing of designated data items without considering data item position in processing queues. Cf. Routine Processing.*
Processing Algorithms	=	*Calibrations and Science algorithms designated by the TMCF for Standard Data Production.*
	=	Scientific Algorithms + Calibration Algorithms
Production-Report	=	Production Schedule + Production Status

Reference-Monitoring-Profile	=	*Expected MODIS instrument engineering parameter levels annotated with limits at which alarm status should be declared.*
Resource-Envelope	=	*Maximum allowable resource consumption levels for the MODIS instrument.*
Retransmission Request	=	*Request for retransmission of data packets that do not meet quality standards.*
Routine Processing	=	*Processing that considers data item position in data processing queues. Cf. Priority Processing.*
Schedule-Data	=	*English language descriptions of planned instrument maneuvers.*
Science-Plan-Coordination	=	*Information exchange between a User requesting special MODIS services and the MODIS Instrument Team Leader. The exchange should culminate in a plan for MODIS Instrument Operation.*
Selected data products	=	*Subsets of standard, near-real-time or specialized data products sent to a user following a pre-established schedule.*
Selected-Real-Time-Science-Data	=	*A subset of MODIS-Science-Data used to monitor MODIS instrument performance.*
Special-Observation-Request	=	*An observation request which requires alteration at previously established observation plans.*
Specialized Data Product	=	*Data products which are considered part of a specific research investigation and are produced for a limited region or time period, or data products which are not accepted by the project as standard items.*
TMCF Observation Request	=	*Request by a TMCF member to execute a MODIS observation request.*

TMCF-Processing-Request	=	*Standard-Processing-Requests + Reprocessing Requests + Data-Base-Inquiry + Selected-Data-Request + Product-Release-Request.*
TMCF Request Response	=	Response to a TMCF-Processing-Request.*
User-Observation-Request	=	*A special instrument measurement request not covered by the current schedule but consistent with general science objectives and the science mission plan.*
User-Processing-Request	=	*Request by a User to generate MODIS-Data-Products not previously available.*
User-Product-Request	=	*Requests that distributed data products be delivered to a User from the MIDACS DADS.*
User-Request	=	*User Product Request + User Observation Request + User Processing Request.*
User-Request-Response	=	*Response to a user's request.*
Science-Data-Request	=	*A request for selected science data for monitoring instrument performance.*
Selected-Science-Data	=	*In response to a request, selected science data will be issued for monitoring instrument performance. Selected data may be of specific channels, time, pockets, etc.*

NOTES FROM CONVERSATION WITH GENE SMITH ON DHC OPERATIONS

Topics Discussed:

1. DHC Data Reception
2. Real-time/Priority Playback/Production Identification
3. DHC Data Distribution

DHC DATA RECEPTION

The DHC receives instrument packet segments from the Virtual Channel Gate Way (VCGW). The VCGW strips away the artifacts associated with the transfer frames, leaving instrument packet segments, which are sent with appended keywords (i.e., error checking information, time, segment counter, etc.) to the DHC. As MODIS packets are sent to the DHC by the VCGW, they may not be contiguous nor complete since various instrument packets were multiplexed onboard by a transfer frame generator that selects a fixed total number of bits for transmission. The beginning segment of each packet will contain the keyword information. The other segments of the packet will be identified by another method (such as time) to associate them with the correct packet. Keywords are used sparingly since they take up room.

REAL-TIME/PRIORITY PLAYBACK/PRODUCTION IDENTIFICATION

The DHC will identify and process data in three ways, real-time, priority playback, and production. Real-time data will undergo error checking, some accounting functions, and then be sent on to the MIDACS. Priority playback data will be processed identically to real-time except that bit order will also be reversed. Production data will undergo the above processes and also management quality checking and accounting, data overlap removal, time ordering, initialization service, and other packet processing. Production data will be sent to the MIDACS only after complete packets have been assembled. Real-time data will be sent to the MIDACS without buffering by the DHC as it receives the data from the VCGW. The DHC will identify real-time packets by reading the packet address code. If the code is recognized as designating real-time, the packet will be immediately sent to the MIDACS. Data can be designated as real-time in two ways: (1) a command can be sent to the MODIS instrument to change the packet address code to match the DHC real-time address code, and (2) the DHC can change its address-interpretation code that determines if the data are real-time.

DHC DATA DISTRIBUTION

The DHC can also combine processing modes so that both real-time and production packets are created. Also, the DHC can select packets to be treated as real-time in a fixed-ratio selection scheme, such as 1 out of each 10 science packets received. The DHC will distribute data to the MIDACS through its own data distribution system. Buffering of data can/will(?) take place here. For real-time data, a minimum time delay is expected (< 1 second). For production work, the data will be distributed after packets have been completely reconstructed. Data can also be saved in the short-term data store at the DHC. Data transmission to MIDACS will be limited by the type of communication line.

COMMENTS

1. Gaps or holes in the real-time data stream may develop because of multiplexing and transfer frame generation. As the DHC receives the data, segments of the a packet may not be received for a significant amount of time (relative to the ICC or CDHF processing and monitoring functions). If the segment is transmitted to the ground at a later time, it will be passed to the MIDACS. If the DHC cannot identify packets segments to place them with the appropriate packet, holes may become permanent.

2. Ancillary data will have its own virtual channel with a data rate of approximately 100 kbps. Some ancillary data will be quality checked, refined, and repaired by the FDF and then sent back to the DHC or directly to the User.

3. There is a possibility that the DHC will have 64 virtual channels for receiving MODIS data; therefore, MODIS may want to request its own virtual channel.

4. Engineering data will be sent to the ICC in real-time by identifying the address code as discussed above.

ACTION ITEMS:

9/16-1: (McKay) Does MIDACS need to request Level 0 data from the DHC routinely, or will the DHC send Level 0 data to MIDACS as soon as it is ready without an explicit request?